

Clinical Practice Guidelines in Practice and Education

Alfred O. Berg, MD, MPH, David Atkins, MD, MPH, William Tierney, MD

Clinical practice guidelines have been a fundamental component of medical practice since one physician first asked another how to manage a patient. A clinical practice guideline is formally defined as a preformed recommendation made for the purpose of influencing a decision about a health intervention.¹ In practice, however, journal articles, editorials, algorithms, care maps, computerized reminders, textbook chapters, and advice from consultants are all clinical practice guidelines in the general form “if x, then y” on some clinical question.

The terminology defining these decision rules is not standard. Physicians will find the terms practice policy, clinical guideline, practice parameter, clinical pathway, standard, algorithm, and many others used synonymously. This article uses the term *clinical practice guideline*.

The recent surge of interest in clinical practice guidelines has several parents. First, medical history is littered with clinical practice guidelines that have been fatally incorrect, leading to interest in methods that promise better validity and reliability. Second, the physician's ability to keep up with the medical literature erodes with each year's burden of (literally) millions of medical articles published worldwide, leading to interest in methods that make sense out of the vast amount of information on a given clinical topic. Third, costly and unexplained variability in medical practice, documented everywhere one looks, leads to interest in developing more accountable approaches for those conditions for which the greatest variations in practice or cost occur (this parent is strongly driven by the shift to more managed care). Fourth, growing demand from patients for greater participation in medical decisions leads to searching for a process in which benefits and harms are linked to outcomes explicitly in terms that patients can understand. Driven by these four parent concerns, methods used to develop clinical practice guidelines have evolved rapidly in recent years. It is important to emphasize, however, that the modern “clinical practice guidelines movement” is too young to have demonstrated success in addressing any of the four concerns, although relevant research attempting to do so is under way in many centers.

Beyond addressing the above concerns, clinical practice guidelines are used for many purposes, some of them

competing. Well-formulated clinical practice guidelines can be used positively not only to guide practice, but also for education, quality assurance and improvement, and cost accountability, ends with which most physicians would agree. On the negative side, guidelines are also used in malpractice actions to justify or attack care provided in specific cases with adverse outcomes, and are used by groups of physicians in attempts to protect clinical turf. Poorly constructed clinical practice guidelines are justifiably attacked when used in any setting, but physicians who assist in developing well-designed evidence-based guidelines must be prepared to find the products used in all kinds of appropriate and inappropriate ways. Attentiveness to the integrity of the process used to generate clinical practice guidelines must be matched with vigilance to guard against inappropriate use.

This article presents an overview of methods used to construct clinical guidelines, discusses an extended example—screening for prostate cancer—and concludes with a review of the use of clinical guidelines in practice and education.

METHODS USED TO CONSTRUCT GUIDELINES

Woolf has outlined four general approaches to developing clinical practice guidelines: informal consensus, formal consensus methods, evidence-based approaches, and explicit approaches.² Under the informal consensus process (or “global subjective judgment”), a group of experts make recommendations based on a subjective assessment of the evidence, with little description of the specific evidence or process used.¹ This approach offers the advantages of simplicity and flexibility, as experts familiar with the topic can reach consensus more quickly and can use clinical and research experience to bridge existing gaps in the evidence. Although informal consensus remains the most common method used to develop guidelines, it is difficult for potential users of such guidelines to judge whether the conclusions are valid or appropriate for a specific population. A consensus process may simply reinforce the biases of the assembled experts, and even when not overtly biased, the guidelines may reflect the narrow perspective of specialists rather than the intended audience.

To guard against some of these problems, more formal consensus methods have been developed, most notably by the Consensus Development Conferences of the National Institutes of Health (NIH), which has produced more than 100 consensus statements over 18 years.^{3,4} To reduce the potential for bias, NIH panels include methodologists, clinicians, and public representatives but exclude persons “identified with strong advocacy positions regarding the topic.” Experts representing a range of opin-

Received from the Department of Family Medicine, University of Washington, Seattle (AOB), Agency for Health Care Policy and Research, Washington, DC (DA), Regenstrief Institute for Health Care, Indianapolis, Ind. (WT).

Address correspondence to Dr. Berg: Department of Family Medicine, University of Washington, Box 356390, Seattle, WA 98195-6390.

ions are invited to present and discuss the evidence at a session open to the public. Other formal approaches to develop expert consensus have been adopted by the American Medical Association Diagnostic and Therapeutic Technology Assessment Program,⁵ the RAND Corporation,⁶ and various health plans.⁷ Because neither formal nor informal consensus methods require that recommendations be explicitly linked to the evidence, it may be difficult to distinguish recommendations based on strong empiric evidence from those based primarily on expert opinion.

In the 1980s, several national and international organizations pioneered efforts to anchor guidelines more directly to the scientific evidence. The Canadian Task Force on the Periodic Health Examination (CTFPHE),⁸ the U.S. Preventive Services Task Force (USPSTF),⁹ and the Clinical Efficacy Assessment Program of the American College of Physicians (in collaboration with the Blue Cross and Blue Shield Association)^{10,11} have each produced a series of guidelines using an evidence-based approach that defines the specific questions to be answered, establishes criteria for including potential evidence, and specifies a systematic process for locating and evaluating relevant evidence. Panels issue recommendations that explicitly reflect the weight of the accumulated evidence. This approach has been credited with improving the reliability and validity of resulting guidelines,¹² and is increasingly being incorporated by other groups. As employed by many groups, however, the evidence-based approach often focuses on a single outcome (e.g., disease-specific mortality) without quantifying other important outcomes.

Eddy has outlined an explicit approach that builds on the evidence-based approach by systematically estimating the effects of interventions on all important health outcomes.¹ He distinguishes between outcome-based approaches, which simply describe the probabilities of various outcomes, and preference-based approaches, which also incorporate patient preferences to determine the best clinical strategy for groups or individuals. Outcomes-based or preference-based methods have been incorporated to varying degrees in a number of guidelines, including some developed by the Agency for Health Care Policy and Research (AHCPR),^{13,14} the American College of Physicians (ACP),¹⁵ and the Office of Technology Assessment.^{16,17}

EXPLICIT AND EVIDENCE-BASED METHODS

In the process of developing evidence-based or outcome-based guidelines, reviewed in a number of recent publications, each step is specified, from selecting topics to crafting precise wording of the guideline.^{12,18–20}

Selecting a Topic and Target Audience

A recent Institute of Medicine report recommended six general criteria to select appropriate topics for guidelines: prevalence of the condition; burden of illness (i.e., morbidity and mortality); costs of treatment; variability in

practice; potential for a guideline to improve health outcomes; and potential for a guideline to reduce costs.¹⁸ Although most organizations loosely follow these recommendations, few have formal processes for selecting topics. Professional societies need also to consider specific needs of their membership, while government agencies may have other political considerations.

It is equally important to specify the target audience and clinical setting for which a guideline is intended. This in turn defines the scope of the guideline, the evidence to be considered, and the composition of the panel. Unless the clinical setting and intended audience are specifically considered, the resulting guideline may be impossible to implement effectively. Many guidelines produced by the Centers for Disease Control and Prevention (CDC), American Cancer Society (ACS), and NIH that deal with primary care issues have paid little attention to critical factors that will determine whether a guideline will be accepted and implemented: cost implications, time or technology required, competing demands of more common problems, and patient preferences.

Panel Composition

Although explicit, evidence-based methods improve the consistency and reliability of resulting guidelines, the composition of the guideline panel can still profoundly influence the process and resulting product. The size and composition of the panel must balance scientific, practical, and political concerns. Guideline panels generally range from 10 to 20 members; within this range, smaller sizes facilitate decision making and decrease costs, but allow less diverse representation. The distribution between specialist-experts and generalists is another important consideration.¹² Recognized experts bring an important perspective to the scientific review process and are an important source of credibility for the resulting guideline. At the same time, experts are more likely to be influenced by their personal involvement with specific treatments or research.¹⁹ Panels can balance this tendency by including individuals with expertise in research methodology, generalist physicians, nonphysician health professionals, and patient or consumer representatives. Multidisciplinary representation, however, makes it harder for all panelists to participate equally in all steps of the process and may complicate reaching consensus. Nonphysician panelists have been routinely and effectively included in AHCPR panels, but the recent USPSTF, CTFPHE, and ACP panels emphasized the role of generalist physicians with methodologic expertise. These panels use extensive peer review of draft guidelines to obtain both expert and multidisciplinary input.

Defining the Causal Pathway

The primary obstacle to evidence-based guidelines remains the paucity of well-designed research for many of

the treatments and technologies in common use. This is a particular problem in assessing the clinical benefits of many screening and diagnostic tests. When direct evidence is not available to link an intervention to an improved health outcome (e.g., cholesterol screening to reduced coronary heart disease), indirect evidence of effectiveness may be provided by studies of intermediate steps in the causal process (e.g., the accuracy of screening tests for identifying patients with high cholesterol, and the effectiveness of cholesterol-lowering treatments in such patients for reducing coronary events).

Battista and Fletcher first outlined a concept of "causal pathways" as an explicit method to describe the use of indirect evidence to establish effectiveness (Figure 1).²¹ An important function of this process is that it identifies both specific beneficial health outcomes and potential adverse consequences at each step of a sequence of interventions. Subjective judgments may still be involved in deciding whether indirect evidence is sufficient (e.g., whether effectiveness of cholesterol-lowering drugs can be inferred for women), but explicitly describing the evidence for each step and outcome allows guideline users to understand and judge the reasoning behind the recommendations.

Literature Retrieval

A systematic review of the literature must be conducted to retrieve relevant evidence for each question identified. Panels may differ in the eligibility criteria they set: some may include any study in which methods can be independently assessed, whereas others may limit themselves to peer-reviewed, published studies. The work of the Cochrane Collaboration may make it easier for panels to include unpublished studies, but these efforts are currently confined to controlled trials. Computerized searches of bibliographic databases such as MEDLINE should be supplemented by reviewing bibliographies of key articles, contacting key experts, and consulting the work of international groups conducting systematic reviews.^{22,23} Government documents and reports that archive important disease-specific information will also be missed in MEDLINE, requiring reviewers to contact government agencies involved in research or program administration for some conditions.

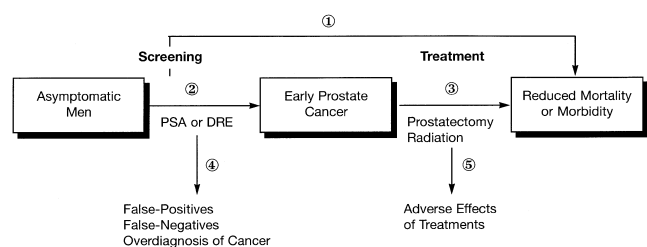


FIGURE 1. Causal pathway of steps linking prostate cancer screening to improved health outcomes.

Literature Review

A critical and often contentious step of the evidence-based process is evaluating the quality of individual studies, including assessment of the internal and external validity of study results. Internal validity is a function of both study design and its implementation. Although certain study designs, (such as randomized controlled trials [RCT's]), are less prone to bias, many factors may compromise the internal validity of an individual trial: inadequate power, improper design or analysis, incomplete follow-up, inappropriate end points, or ineffective implementation of the intervention. Confounding, selection bias, and measurement error are important considerations for observational studies. Standardized methods for assessing individual studies have been described.^{1,22,24} Equally important is whether the results of an individual study are relevant to the general population (i.e., external validity). Unrepresentative study populations or clinical settings and impractical interventions may compromise the external validity of an otherwise well-conducted clinical trial. Even when a study provides good evidence of efficacy (i.e., results under ideal conditions), guideline developers need to consider that the effectiveness of an intervention (i.e., results under typical conditions) may be substantially lower when implemented in the real world.

Panels will often need to turn to expert opinion to address important questions for which definitive evidence is not available. Both informal methods and formal methods (e.g., the Delphi technique) can be used, but care should be taken to solicit a broad range of opinion and to clearly distinguish recommendations supported by the evidence from those based primarily on opinion.

Developing a Balance Sheet of Benefits and Harms

Summary estimates of the effects of alternative treatment strategies can be represented as a balance sheet depicting discrete benefits (e.g., reduced mortality and morbidity, improved quality of life) and harms (fatal and nonfatal complications or side effects) for each strategy. When patient preferences can be estimated for specific outcomes, outcomes can be converted into a common denominator such as quality-adjusted life years to calculate the net effect (benefits minus harms) of a certain strategy. When treatments have clear effects on mortality with limited adverse effects (e.g., angiotensin-converting enzyme inhibitors for congestive heart failure), detailed balance sheets may not be necessary. However, when interventions primarily improve symptoms but involve important risks (e.g., surgery for benign prostatic hyperplasia), such balance sheets can help illustrate the tradeoffs involved. Cost-effectiveness, as a cost per life year or quality-adjusted life year, can be modeled. Unfortunately, the complexity of such models and the numerous assumptions that must be made to estimate parameters for them, may make their

conclusions less credible to the average clinician, especially when results contradict prevailing practice.

Linking Evidence to Recommendations

The hallmark of the explicit, evidence-based approach is that each recommendation directly reflects the strength of underlying evidence. Different groups have used different hierarchies to depict this link. The USPSTF and CTF-PHE use separate hierarchies to describe both the level of the evidence (I–III) and the strength of the overall recommendation (A–E), whereas AHCPR panels have used simpler A–B–C categories to distinguish recommendations based on definitive trial evidence, those based on weaker evidence, and those based on expert opinion.

Levels of Flexibility

Guideline panels inevitably struggle to balance their desire to create guidelines that are forthright and unambiguous, with the recognition that no guideline can supplant the role of clinical judgment about the individual patient. Often, guidelines seek to convey these distinctions with subtle language distinctions (“X may be useful”). As guidelines are increasingly being used to measure quality or guide reimbursement, they will need to distinguish those measures for which there is clear and compelling evidence of important benefits (e.g., childhood vaccinations), measures that should be encouraged but not required (e.g., sigmoidoscopy), and interventions for which the best strategy depends largely on clinician judgment and patient preference (e.g., hormone replacement therapy). The American Academy of Family Physicians has proposed such an approach in adapting the recommendations of the USPSTF, classifying services as “standards,” “guidelines,” or “options.”

Review

The final component necessary to ensure the reliability of clinical guidelines is peer review from a range of outside reviewers, including content experts, representatives of professional societies, government organizations and consumer groups, and potential guideline users. It is now important that the latter group include representatives of managed care and insurers along with practicing clinicians, as these groups may be critical components of guideline implementation. Although it is useful and important to solicit comments from a wide range of reviewers, it is also important to recognize that the perspective and motivation of many reviewers may differ from those of the intended audience. For example, subspecialists are important for addressing scientific issues but most likely to have different priorities than primary care clinicians.

Example

Recent guidelines on prostate cancer screening issued by the ACP¹⁵ and by the USPSTF⁹ illustrate two approaches to explicit, evidence-based guidelines and the strengths and weaknesses of each approach. Figure 1 depicts a causal pathway describing the steps linking screening with prostate-specific antigen (PSA) to improved health outcomes. At present, there is no direct evidence that men screened for prostate cancer have lower mortality or morbidity from prostate cancer compared with unscreened men (step 1); a large, multicenter randomized trial of PSA screening is under way to address this issue. To establish support for screening in the absence of such evidence, one must demonstrate that screening can improve early detection of prostate cancer (step 2), that treatment of asymptomatic prostate cancer is more effective than treatment of clinically apparent cancer (step 3), and that these benefits justify the potential adverse effects of screening (step 4) and treatment (step 5). Most proponents of screening base their argument on the growing evidence for step 2 with an assumption that early treatment can improve outcomes, based primarily on prostatectomy series demonstrating favorable outcomes for patients with localized prostate cancer compared with those who have advanced disease.

Evidence-based groups have recognized numerous problems in using such uncontrolled data to infer a benefit of early detection: lead-time bias, length bias, and selection bias all could result in apparent survival benefit even if treatments were ineffective; mortality results ignore important consequences of treatment on quality-of-life; and experience with cancers detected in the pre-PSA era may not apply to the small, indolent cancers now being detected with widespread screening. The lack of evidence that early detection was beneficial, and concern that risks of screening and aggressive therapies could actually exceed benefits, led the USPSTF⁹ and CTFPHE⁸ to discourage routine screening in asymptomatic men. Such a recommendation reflected the underlying philosophy of these groups that, when proposing interventions to asymptomatic persons, clinicians should be guided by a principle of “first do no harm.”

These recommendations can frustrate clinicians who are concerned about the toll of prostate cancer and see a plausible (if not yet proven) potential for early detection to reduce cancer deaths. The assessment of prostate cancer screening done by the Office of Technology Assessment¹⁹ and the recently released guideline developed by the ACP from a similar analysis¹⁵ used an outcome-based model to explore the range of possible benefits and risks of screening in different populations. The models used available evidence to estimate the accuracy of PSA, prevalence and prognosis of cancers of various grades, and the costs and complications of screening and prostatectomy (see Table 1); because data on treatment efficacy are lacking, they used assumptions favorable to screening (100% effectiveness

for organ-confined cancers) and sensitivity analysis to examine how screening decisions might differ if treatments were less effective. As depicted in Table 1, even if treatments are 100% effective, the benefits of screening men over 70 years of age (cancer deaths prevented) are offset by the large number of men who suffer excess morbidity from treatment. Under the same assumptions, screening younger men (aged 50–70 years) offered a more favorable balance of risks and benefits (although the absolute benefit remains small) and could be reasonably cost-effective.¹⁵ Screening men over 70 is cost-ineffective under the most optimistic of assumptions (>\$65,000 per life year saved) and could result in a net decrease in quality of life if treatments are less than optimally effective or if lower disease-progression rates are assumed.^{15,19} Although these analyses require a number of assumptions, they make clear that the value of screening depends critically on how effective therapies are and on how likely untreated cancers are to progress in a patient's lifetime. Although consistent with the fundamental uncertainty of the evidence, the ACP guideline¹⁵ provides room for clinicians and patients to reach their own decision about dealing with an intervention of possible but unproven benefit: it recommends against screening average-risk men before age 50 or after age 70, but leaves screening optional for men aged 50 to 70 years. If screening is offered, however, it should be accompanied by a careful explanation of its potential benefits and risks and the underlying uncertainties about effectiveness.

USE OF GUIDELINES IN PRACTICE

Clinical practice guidelines have always been at the heart of medical practice. The current national dialogue

Table 1. Balance Sheet of Benefits and Harms of One-Time PSA Screening in Men Age 70 (per 100,000 men screened)*

Benefits and Harms [†]	N	Life-Years Saved	Life-Years Improved
Estimated benefits			
Cancer deaths prevented	449	3,019	
Metastatic cases prevented	1,019		829
Estimated harms			
Positive PSA or digital rectal exam	34,000		
Biopsies	23,460		
Radical prostatectomy	3,360		
Deaths [‡]	22	–245	
Incontinent	301		–3,229
Impotent	1,569		–16,908
Incontinent + impotent	467		–5,050
Total		2,774	–24,358

*Adapted from Office of Technology Assessment.¹⁷

[†]Net benefits/harms per patient screened: 10 days longer life, with 89 days worse quality of life. Net benefits/harms per patient treated: 301 days longer life, with 2,646 days worse quality of life.

[‡]Includes one death from urosepsis due to biopsy.

on guidelines is driven not so much by the question of whether there should be guidelines, but by how to see that they are followed. This emphasis on effective implementation is due to the unacceptably high cost of health care and has the two goals of avoiding paying for unnecessary (or unnecessarily high-cost) health care, and making sure that, when costs are cut, the quality of health care does not suffer.

Implementing guidelines can be cost-saving, especially for those that are specifically written to deal with the costly aspects of health care (e.g., “negative drug detailing” aimed at reducing the use of unnecessarily costly drugs). However, balanced practice guidelines should deal with errors of omission as well as commission. Correcting errors of omission (e.g., not providing indicated preventive care, or not treating patients with left ventricular dysfunction with a converting enzyme inhibitor) may, at least in the short run, be more costly. In such cases, the practitioners must define as high-quality care actions that are cost-effective, if not cost-saving. They can then use guidelines to encourage (or enforce) such care. Providing the best care, even if the clinical benefits are in the future, is the “good” extracted at the “cost” of using practice guidelines. Finally, one must not exclude from calculations of the cost-effectiveness of guidelines the costs of developing guidelines (costs = personnel time) and implementing them (costs = data extraction + data management + provision of interventions of feedback to providers).

The aspects of health care likely to reap the most benefit from the application of practice guidelines are those for which there is demonstrable variation in care provided, costs, or outcomes. One might begin with the most serious and costly conditions for which compliance with well-constructed guideline protocols is suboptimal. For example, guideline efforts would be better focused on thrombolytic care for acute myocardial infarctions in the emergency department or influenza vaccination of elderly outpatients than on screening for hypercholesterolemia among asymptomatic elders.

The first controlled trials of guideline implementation were reminder studies in which the most common and effective reminders were triggered by simple demographic data (age and gender) and compliance was monitored by noting whether an appropriate order was written, a test result was obtained, or a drug was prescribed.^{25–28} Even simple guidelines may prove difficult and time-consuming to design and implement at first; but with experience and improved data systems, more complex guidelines will be easier to accomplish. The most sophisticated guideline systems currently are those in which orders for specific testing and therapy are suggested via algorithms built into physicians' order-writing workstations.²⁹ Early studies show that, when physicians intend to comply with such suggestions, compliance is enhanced.³⁰ When they believe the suggestions are inappropriate, they do not comply (which is reassuring).³¹

Physicians and other health care providers are not ac-

customed to having their care audited. A common response to the promulgation and enforcement of guidelines is that they will lead to “cookbook medicine.” Of course, blindly following general rules will obscure valid individual differences and result in poor care. Guidelines should only apply to between 60% and 95% of relevant cases.³² Yet physicians often further simplify decision-making processes by invoking rules of thumb termed heuristics.³³ Such rules (a form of guideline) are one way of dealing with information overload by simplifying complex rules and information matrices into a smaller number of overriding “truths.” Occam’s razor is a familiar heuristic: choose the simplest hypothesis that explains the most findings.³⁴

Two keys are necessary in order to apply guidelines in clinical practice. First, the providers within a practice must believe that the guidelines are appropriate and intend to follow them. Second, the practice must have access to sufficient data and sufficiently specified rules.

Developing Consensus

Intention to comply with guidelines is the strongest predictor of compliance.²⁵ Conversely, when physicians do not intend to comply with guidelines, they will not comply despite being reminded to do so on a case-by-case basis.³¹

Physician confidence in the correctness of a guideline depends on the methods used to construct the guideline and who does the work. Examples of evidence-based guidelines are those from the USPSTF,⁹ the CTFPHE,⁸ and the AHCPR. Guidelines can also be established by the consensus of medical experts. Examples include guidelines for the treatment of heart disease from the American College of Cardiology³⁵ and the long series of consensus conferences held by the NIH.⁴ Internists have more faith in specialty societies than in government or regulatory agencies.³⁶ This is likely due to the belief that physicians are motivated to provide the highest-quality care, while the motives of other organizations may be suspect. Internists fear that the goal of guidelines promulgated from nonprofessional organizations is mainly to reduce the cost of care.³⁶

Implementing Guidelines in Everyday Practice

Many attempts at creating clinical practice guidelines are no more than a rendering of medical textbook advice into algorithms, which often wind up on bookshelves and not incorporated into medical practice. To make operational and enforce clinical practice guidelines, one must have sufficient data to identify patients who might be covered by the guideline, specify what should be done for the majority of eligible patients, assess what was actually done, and monitor patient outcomes. For example, the AHCPR heart failure guidelines define the eligible patient as one who has a left ventricular ejection fraction (measured by two-dimensional echocardiography) of less than 35% to 40%.³⁷ To avoid debate on whether particular pa-

tients are covered by the guideline, an explicit cutoff value for this parameter must be established.^{38,39} But what if there is no echocardiogram result for a patient, or the echocardiogram report does not routinely quantify ejection fraction? Can other data (e.g., diagnoses or other cardiac imaging studies) be substituted? Local translation and implementation of guidelines includes not only adopting guidelines that the providers intend to follow but also modifying those guidelines to make them compatible with the exigencies of each practice.

The goal of clinical practice guidelines is to improve the quality and cost-efficiency of health care. If this is to occur, then monitoring of patient care (i.e., gathering data) is a must. There are only two sources for such data: the paper chart and electronic medical record systems. Extracting information from paper records is time-consuming, expensive, and error-prone. Moreover, the paper record usually contains information from only one site of care (e.g., a physician’s office or a hospital). Extracting the needed information from existing electronic medical record systems is less expensive and more accurate and will eventually be the source of most of the data needed to operationalize guidelines.⁴⁰ Although such systems are increasingly prevalent, they are idiosyncratic and usually only contain part of the needed data (such as laboratory test results or pharmacy records). Most electronic medical record systems capable of monitoring guidelines should also be capable of improving compliance with them by reminding physicians when their patients are eligible for guideline-directed care. Physicians respond favorably to such reminders or “ticklers,”^{26,27,41} but the effect is greatest when they intend to comply with the underlying guideline rules that drive the reminders.²⁵

It is critical to recognize that not all guideline rules will be followed, and that this “noncompliance” will often be appropriate.⁴² In fact, medicolegal risk will be lessened if systems to invoke guidelines offer physicians a means for permanently recording their reasons for noncompliance. For example, a woman eligible for a Pap smear might refuse it or be menstruating. A man with metastatic lung cancer would not be a candidate for fecal occult blood screening. An angiotensin-converting enzyme inhibitor may have been used elsewhere for heart failure and have given the patient an intractable cough, or a patient might be acutely ill so that guideline-directed care should be delayed.

The perception that guidelines reduce the role of clinical judgment poses a barrier to wider acceptance by clinicians. Treatment recommendations in guidelines generally emphasize evidence of effectiveness, but clinicians must also consider other factors such as patient preferences, costs, competing health priorities, and the magnitude of the benefit when dealing with individual patients. Moreover, individual patients might decide against an effective intervention because the small absolute benefit does not seem to justify the inconvenience, costs, or risks (e.g., sigmoidoscopy screening). Conversely, clinicians

and patients may favor some unproven interventions because they value the possible benefits more than they fear the potential risks (e.g., prostate cancer screening). Although comprehensive guidelines can try to take factors such as costs and patient preferences into account, no guideline can anticipate their relative importance for every case. Thus, recommendations that would be beneficial on average may not be for everyone.

As tools for quality improvement, guidelines may be most useful in defining goals against which we can measure the care of a population of patients rather than of every member. This has important implications for how one implements and monitors guidelines. Rather than scrutinizing whether every patient was managed by protocol, guidelines can be used to define a few discrete performance measures for providers and systems of care (e.g., percentage of eligible patients receiving mammography within the previous 1 to 2 years). Performance data can be fed back to individuals and systems, and the data used to plan appropriate targets for quality improvement.

Guidelines and Managed Care

Managed care organizations will be particularly interested in implementing practice guidelines to optimize the cost-effectiveness of care and define appropriate care, guarding against the times when a patient's health may be in conflict with capitated practice income. This means, however, that such organizations must commit sufficient resources to choose appropriate guidelines (i.e., they speak to a relevant clinical problem and are sufficiently specified); modify them to reflect local standards of practice and information resources; obtain consensus from the providers in that practice; develop the protocols for implementing them; assess compliance; deal with non-compliance; monitor patient outcomes (including clinical processes and outcomes of care as well as patients' health status and satisfaction); and continually monitor the relevant literature and update the guidelines as needed. These tasks are far from trivial and can themselves be costly, so managed care organizations should deal first with clinical issues when both quality of care and costs of care are at issue simultaneously.

Other Implementation Concerns

Although a rigorous methodology can improve the scientific validity of guidelines, other factors may be more important in determining whether a guideline can be implemented effectively to improve clinical practice. Over the past 5 years, the AHCPR sponsored 19 comprehensive clinical practice guidelines, produced by multidisciplinary expert panels using explicit, evidence-based methods. Although the guidelines were generally praised for their rigor and objectivity, they proved expensive and time-consuming and became a target for special interests who felt threatened by specific recommendations. Moreover, feder-

ally sponsored panels were not as well positioned as professional societies, managed care organizations, and local institutions to anticipate the critical issues for implementing specific recommendations. These factors motivated recently announced changes in AHCPR's guideline activity. Rather than developing detailed clinical guidelines, AHCPR has announced that it will work in partnership with public and private organizations who have identified important topics in need of guidelines. AHCPR would contract with outside centers to summarize the scientific evidence as a background report including, when appropriate, decision analysis, cost analysis, and cost-effectiveness analysis, but the partners will be responsible for turning the evidence into appropriate clinical recommendations for their audience. At the same time, AHCPR is working with several organizations to develop a national clearinghouse of clinical guidelines. This clearinghouse would provide easy access to wide range of guidelines, produce a standardized abstract for each guideline (including topic, authors, target audience, methodology, recommendations, etc.), and provide a synthesis explaining areas of agreement and disagreement among guidelines on a given topic. Finally, AHCPR plans to develop new research initiatives to study the most effective ways to implement guidelines and other quality-improvement tools.

USE OF GUIDELINES IN EDUCATION

Practice guidelines of one sort or another have been at the core of clinical teaching for generations, representing attempts by teachers to reduce complex clinical questions to simple rules for diagnosis and treatment that trainees can learn and apply. Educationally, current guidelines can be seen as an attempt to produce "virtual" textbooks in algorithmic form. We are aware of no data on the effects of guideline implementation on educational achievement; that is, no clinical trials have been conducted showing that an educational program based on an evidence-based guideline approach produces more competent physicians than one based on usual educational methods. Our discussion here, then, must be seen as preliminary and speculative.

For student and resident education, the implications of the clinical practice guidelines movement using the newer evidence-based methods are potentially profound. First, most clinical education is based on an expert or preceptorial model, conceptually far removed from the explicit and evidence-based model used by high-quality clinical practice guidelines. Second, gaps in evidence uncovered in the process of developing an explicit and evidence-based guideline are ubiquitous, raising uncomfortable questions about the validity of teaching in those situations in which the evidence to reach a scientific conclusion is simply unavailable. Third, although many teaching programs do a good job of training in critical appraisal of original research, the clinical practice guidelines movement suggests that the theory that an *individ-*

ual study should ever be used to guide practice (the "silver bullet" theory of medical literature) is untenable. Educators would be better advised to train students and residents how to evaluate clinical practice guidelines, using rigorous criteria to judge the quality of the process and outcome. Fourth, there is an educational hazard in teaching the use of clinical guidelines (even excellent ones) rigidly and uncritically. A system using guidelines in education must teach not only how to choose a guideline, but when to apply it, to whom, and, importantly, when its use would be inappropriate.

Failure to equip trainees with the skills necessary to evaluate and apply clinical practice guidelines will leave them with a serious disability as guidelines become the norm for practice. Students and residents untrained in the new methods will not be able to intelligently negotiate their way; they will tend to be overly deferential to authority, and unable to critically evaluate clinical practice guidelines promulgated and advocated by various health care organizations. They will lose autonomy, forced to follow poorly constructed guidelines that subject their patients to substandard care without a strong scientific base.

Every physician teaching students and residents should become familiar with methods used to construct and evaluate explicit and evidence-based clinical practice guidelines. Methods to train faculty in clinical practice guidelines and evidence-based medicine are still very much in development.

Specific curricular strategies might include seminars, readings, journal-club-like discussions of recently released policies, selective use of clinical practice guidelines in practice, involvement in quality assurance and improvement activities using guidelines, and so on. The important steps are the same as for any other curricular intervention: design, implementation, and evaluation. Part of curricular intervention will be persuading clinical colleagues in other fields to take the movement seriously and not sabotage efforts to teach the material. There is residual hostility among physicians about clinical practice guidelines in general, and toward evidence-based ones in particular. The reasons for this are many, including distaste for so-called cookbook medicine, fear of malpractice, a desire to fit into community norms, and others. These are all valid concerns, but each has a valid response, and each must be overcome if guidelines are to achieve their potential usefulness. Physicians would do well to transform hostility into healthy skepticism, including careful assessment of guidelines and their relevance to teaching and practice.

SOURCES OF GUIDELINES

Some sources of clinical practice guidelines have been mentioned in the text or in cited references. Table 2 provides details regarding where some of the more commonly used guidelines can be found. The authors recommend that practitioners and educators focus on guide-

Table 2. Sources of Information About Clinical Practice Guidelines

1. AMA Directory of Practice Parameters. Chicago, Ill: American Medical Association; 1996. Publication OP270396.
This directory collects in one place most clinical practice policies generated from national organizations and governmental agencies. It is the best available directory, but does not attempt to critically review any of the collected policies.
2. Agency for Health Care Policy and Research
Call 1-800-358-9295 for free individual copies of quick reference guides on their 20 released guidelines. Ordering information for the full reports is also available. Full text is available on the National Library of Medicine's Internet site: <http://text.nlm.nih.gov/ftsr/pick?collect=ahcpr&ftsrK=0>.
3. U.S. Preventive Services Task Force. Guide to Clinical Preventive Services. 2nd ed.
The Task Force recently published its evidence-based reviews of nearly 200 preventive interventions. It is available from two commercial publishers (Williams and Wilkins, Baltimore, Md.; International Medical Publishing, McLean, Va.); from the U.S. Government Printing Office (202-512-1800; stock no. 017-00525-8); and on the Internet at two sites:
<http://odphp.osophs.dhhs.gov/pubs/guidecps>
<http://text.nlm.nih.gov/ftsr/pick?collect=cps&ftsrK=0>
4. Canadian Task Force on the Periodic Health Examination. The Canadian Guide to Clinical Preventive Health Care. Ottawa, Ont: Health Canada; 1994.
5. American College of Physicians
The ACP, through its Clinical Efficacy Assessment Project, has produced a long series of well-formulated clinical practice guidelines. Many have been published in the Annals of Internal Medicine. Individual copies can be ordered at 1-800-523-1546.

lines with excellent evidence-based provenance, with special attention to those from the AHCPR, the U.S. and Canadian Prevention Task Forces, and the ACP.

REFERENCES

1. Eddy DM. A Manual for Assessing Health Practices and Designing Practice Policies: The Explicit Approach. Philadelphia, Pa: American College of Physicians; 1992.
2. Woolf SH. Practice guidelines, a new reality in medicine, II: methods for developing guidelines. Arch Intern Med. 1992;152:946-52.
3. Ferguson JH. NIH Consensus Conferences: dissemination and impact. Ann NY Acad Sci. 1993;703:180-99.
4. Ferguson JH. The NIH Consensus Development Program. Int J Technol Assess Health Care. 1996;12:460-74.
5. American Medical Association. Diagnostic and Therapeutic Technology Assessment. DATTA Procedures. Chicago, Ill: American Medical Association; April 1992.
6. Brook RH, Chassin MR, Fink A, et al. A method for the assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care. 1986;2:53-63.
7. Gottlieb LK, Sokol HN, Murrey KO, et al. Algorithm-based clinical quality improvement: clinical guidelines and continuous quality improvement. HMO Pract. 1993;6:5-12.
8. Canadian Task Force on the Periodic Health Examination. The Ca-

- nadian Guide to Clinical Preventive Health Care. Ottawa, Ont: Canada Communication Group; 1994.
9. U.S. Preventive Services Task Force. Guide to Clinical Preventive Services. 2nd ed. Baltimore, Md: Williams and Wilkins; 1996.
 10. Eddy DM, ed. Common Screening Tests. Philadelphia, Pa: American College of Physicians; 1991.
 11. Sox HC Jr., ed. Common Diagnostic Tests: Use and Interpretation. 2nd ed. Philadelphia, Pa: American College of Physicians; 1990.
 12. Institute of Medicine, Committee on Clinical Practice Guidelines. Field MJ, Lohr KN, eds. Guidelines for Clinical Practice: From Development to Use. Washington, DC: National Academy Press; 1992.
 13. Woolf SH. ACHPR Interim Manual for Clinical Practice Guideline Development. Rockville, Md: Agency for Health Care Policy and Research; 1991. AHCPR publication No. 91-0018.
 14. McConnell JD, Barry MJ, Bruskewitz RC, et al. Benign Prostatic Hyperplasia: Diagnosis and Treatment. Clinical Practice Guideline No. 8. Rockville, Md: Agency for Health Care Policy and Research; February 1994. AHCPR publication No. 94-0582.
 15. Coley CM, Barry MJ, Fleming C, Fahs MC, Mulley AG. Early detection of prostate cancer. Part II: Estimating the risks, benefits, and costs. *Ann Intern Med.* 1997;126:468-79.
 16. Office of Technology Assessment. Effectiveness and Costs of Osteoporosis Screening and Hormone Replacement Therapy, Vol II: Evidence on Benefits, Risks, and Costs. Washington, DC: U.S. Congress; August 1995. Publication No. OTA-BP-H-144.
 17. Office of Technology Assessment. Costs and Effectiveness of Prostate Cancer Screening in Elderly Men. Washington, DC: U.S. Congress; May 1995. Publication No. OTA-BP-H-145.
 18. Institute of Medicine, Committee on Methods for Setting Priorities for Guidelines Development. Field MJ, ed. Setting Priorities for Clinical Practice Guidelines. Washington, DC: National Academy Press; 1995.
 19. Office of Technology Assessment. Identifying Health Technologies That Work: Searching for Evidence. Washington, DC: U.S. Congress; September 1994. Publication No. OTA-H-608.
 20. Woolf SH, DiGiuseppi CG, Atkins D, et al. Developing evidence-based clinical practice guidelines: lessons learned by the U.S. Preventive Services Task Force. *Annu Rev Public Health.* 1996;17:511-38.
 21. Battista RN, Fletcher SW. Making recommendations on preventive practices: methodologic issues. *Am J Prev Med.* 1988;4(suppl): 53-67.
 22. Bero L, Rennie D. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA.* 1995;274:1935-8.
 23. NHS Centre for Reviews and Dissemination. Undertaking Systematic Reviews of Research on Effectiveness: CRD Guidelines for Those Carrying Out or Commissioning Reviews. CRD Report No. 4, January 1996. York, England: University of York; 1996.
 24. Oxman AD, ed. The Cochrane Collaboration Handbook. Section VI: Preparing and Maintaining Systematic Reviews. Oxford, England: Cochrane Collaboration; 1994.
 25. McDonald CJ, Hui SL, Smith DM, et al. Reminders to physicians from an introspective computer medical record. A two-year randomized trial. *Ann Intern Med.* 1984;100:130-8.
 26. Barnett GO, Winickoff RN, Morgan MM, Zielstorff RD. A computer-based monitoring system for follow-up of elevated blood pressure. *Med Care.* 1983;21:400-9.
 27. McPhee SJ, Bird JA, Fordham D, Rodnick JE, Osborn EH. Promoting cancer prevention activities by primary care physicians. Results of a randomized, controlled trial. *JAMA.* 1991;266:538-44.
 28. Tierney WM, Hui SL, McDonald CJ. Delayed feedback of physician performance versus immediate reminders to perform preventive care: effects of physician compliance. *Med Care.* 1986;24:659-65.
 29. Sittig DF, Stead WW. Computer-based physician order entry: the state of the art. *J Am Med Inform Assoc.* 1994;1:108-23.
 30. Overhage JM, Tierney WM, McDonald CJ. Computer-assisted order-writing improves compliance with ordering guidelines. *Clin Res.* 1993;41:716A. Abstract.
 31. Overhage JM, Tierney WM, McDonald CJ. Computer reminders to implement preventive care guidelines for hospitalized patients. *Arch Intern Med.* 1996;156:1551-6.
 32. Eddy DM. Designing a practice policy: standards, guidelines, and options. *JAMA.* 1990;263:3077-82.
 33. McDonald CJ. Medical heuristics: the silent adjudicators of clinical practice. *Ann Intern Med.* 1996;124:56-62.
 34. Sapira JD. On violating Occam's razor. *South Med J.* 1991; 84:766.
 35. Guidelines for the evaluation and management of heart failure. Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Evaluation and Management of Heart Failure). *J Am Coll Cardiol.* 1995;26:1376-98.
 36. Tunis SR, Hayward RS, Wilson MC, et al. Internists' attitudes about clinical practice guidelines. *Ann Intern Med.* 1994;120:956-63.
 37. Agency for Health Care Policy and Research. Heart Failure: Evaluation and Care of Patients with Left-Ventricular Systolic Dysfunction. Washington, DC: U.S. Department of Health and Human Services; 1994. AHCPR publication No. 94-0612.
 38. Tierney WM, Overhage JM, Takesue BY, et al. Computerizing guidelines to improve care and patient outcomes: the example of heart failure. *J Am Med Inform Assoc.* 1995;2:316-22.
 39. McDonald CJ, Overhage JM. Guidelines you can follow and can trust: an ideal and an example. *JAMA.* 1994;271:872-3.
 40. Institute of Medicine, Committee on Improving the Medical Record. The Computer-Based Patient Record: An Essential Technology for Health Care. Washington, DC: National Academy Press; 1991.
 41. Rind DM, Safran C, Phillips RS, et al. Effect of computer-based alerts on the treatment and outcomes of hospitalized patients. *Arch Intern Med.* 1994;154:1511-7.
 42. Litzelman DK, Tierney WM. Physicians' reasons for failing to comply with computerized preventive care guidelines. *J Gen Intern Med.* 1996;11:497-9.